EVALUATION & MONITORING OF CISE’s PATHWAYS TO REVITALIZED UNDERGRADUATE COMPUTING EDUCATION (CPATH) PROGRAM

NSF MOBIS Contract no. GSIOF0554N
Requisition DACS-08D1530

Logic Models for the CPATH Project

Prepared for:
Harriet G. Taylor
CISE
National Science Foundation
4201 Wilson Boulevard, Room 1270
Arlington, Virginia 22230

Kay Marie Rison, Contract Specialist
National Science Foundation
Division of Acquisition and Cooperative Support (DACS)
4201 Wilson Blvd., Suite 475
Arlington, VA 22230
703-292-2152
mrison@nsf.gov

Prepared by:
SRI International
Science and Technology Policy Program
1100 Wilson Blvd., Suite 2800
Arlington, VA 22209-3915
Contact: Caroline S. Wagner
703-247-8478

SRI Project Number P18680
UNDERSTANDING THE LOGIC MODELS

This report provides logic models and a glossary of evaluation terms to support CISE’s Pathways to Revitalized Undergraduate Computing Education (CPATH) Program. Providing evaluation input and support to CPATH is a complex and challenging process. To help inform the project, SRI has developed logic models and a glossary of evaluation terms to guide the evaluation process and discussion. These logic models are the result of in-depth discussions with NSF staff. They drew ideas from a CPATH workshop held in November 2008. Following NSF’s approval of this document, these items will be placed on the project website at www.cpathmonitor.org to provide support to the CPATH projects.

Logic models are a common feature of program evaluation techniques. As a general rule, a logic model is schematic flow chart that displays a program’s conceptual design. It serves to guide the plan for evaluation design, data collection, and analysis. Logic models are generally organized to show the flow of events from left to right. They generally contain boxes that include lists of (1) program inputs (e.g., specifications, funding, proposals); (2) program activities (e.g., review process and results); (3) activities of other parties (e.g., research conducted); (4) intermediate outputs (publications, new proposals); and (5) impacts and outcomes (e.g., advances in concepts or instruments, societal effects), (6) external factors. Arrows show the direction of influences of each component on the others. Feedback loops are also shown (e.g., from outputs and outcomes back to program design and funding levels).

By depicting “a plausible and sensible map of how the program will work [to address] identified problems” ¹ a logic model provides opportunities for discourse among evaluators and program managers about the program’s objectives, activities, expectations, and limitations. By exploring the assumptions and expectations underlying a program, the logic model broadens the evaluators’ understanding of the program, and develops shared concepts between the SRI staff, the NSF staff, and the Program awardees who are conducting sponsored research. And by mixing explicit and implicit influences into the model, and enumerating direct and indirect environmental factors, those associated with the project are likely to choose those tools and indicators that will provide robust and effective program evaluation.

The logic models and glossary of terms developed for this project represent the first step in creating supporting documents for the SRI input to the CPATH project. To date, the task has involved first, studying the documents related to CPATH, followed by a series of discussions with NSF staff as we created these logic models to guide the development of the monitoring tool and project evaluation protocol. The logic models illustrate the influences that led to the creation of the CPATH program, NSF’s actions, and the goals and expected outcomes of the categories of projects being conducted under the program. Measures are selected on the basis of the extent to which they will shed light on the progress of the program as a whole, or on progress within sub-sets of the program. The

---

measures are drawn from evaluation literature as ones that are most relevant to the goals being pursued by CPATH. Many of these measures will be built into the program monitoring activities that SRI will build during the course of the Program.

The logic models contained herein are presented at two levels: 1) the program level, and 2) the track and sub-track level. Within both of these types of models, there are subsets of models. The program level models are differentiated at the level of measures relevant to institutional transformation or community building activities. These program level models are:

- Model 1A – Annual Program Evaluation Logic Model for Institutional Transformation Aspects of CPATH
- Model 1B- Annual Program Evaluation Logic Model for Community Building Aspects of CPATH.

The award level models are represented by three different types, one for each type of CPATH project: 1) institutional transformation; 2) community building; and 3) conceptual development and planning. Within each of these types, the models are further differentiated into boxes at four levels: 1) operations, 2) outputs, 3) outcomes, and 4) measures. Some measures listed at this final level will be the same across two or three of the models, but they are repeated in the relevant logic model. These award-level models are:

- Model 2 - Award-level Evaluation Models for the Three Types of Programs: Transforming Institutions
  A. Community Building
  B. Conceptual Development and Planning
  C. Institutional Transformation

The sections below explain the boxes within the models.

Models 1A and 1B

1. Precedents: External Influences, National and NSF Context

The logic models should be read beginning at the top, left-hand side of the page. The models all begin with a list of external influences on computing education, and the context within which the nation and NSF find themselves with regard to computer science education. The genesis of the program grew from widespread concern about the state of computer science education and questions as to why enrollment at the undergraduate level was diminishing; this is reflected in the first two boxes. (The far left-hand column of the models is the same in all three types of models, including the box listing NSF’s initial reaction to the issues and problems that appeared to be emerging from the national context.) At the highest level of aggregation, we assume that the influences on NSF will be the same, therefore these boxes are identical in each of the logic models.

The external influences were derived from NSF materials related to CPATH, from conversations with NSF staff, from materials in the proposals offered to NSF for sponsored research, as well as from materials available in the media related to computer
science education. These were distilled into the issues and questions that appeared most relevant to the CPATH Program. It is considered that the content of these boxes will change over time, but that the factors will change slowly compared to other parts of the logic model; therefore the color of the text is lighter than other parts of the model. (One of the points of evaluation will be to consider the extent to which CPATH awards help to change the external influences and national context within which NSF operates.)

Precedents: NSF CISE Actions

In considering the response that NSF should make to what was widely perceived as a crisis in computing education, NSF had a number of tools to put into action. The first is was to engage the NSF Computer, Information Science and Engineering (CISE) Directorate to design a program and put it into operation. The box on the far left displaying the NSF context lists the actions that NSF took as it began the CPATH Program, including hosting regional workshops, considering different types of projects, writing and fielding a solicitation, and choosing and support specific projects. This process takes place each year, so the contents of the box marked “NSF CISE Actions” can be expected to change each year as the context changes.

2. Internal NSF Actions: Solicitation and Award Characteristics, Project-level Administration Conditions, SRI Monitoring

Model 1 shows a box with the award characteristics of the projects that have been selected for funding by CPATH. This includes the three tracks: 1) community building, 2) institutional transformation, 3) conceptual development and planning, and we added a fourth overall category in which NSF chose to include geographic spread and institutional diversity (colleges, historically minority institutions) in the mix of choosing institutions that received support under the program. A subset of these characteristics is whether the award goes to single institutions, or a coalition of institutions that have joined together to propose a project under CPATH. These can include multiple academic institutions, cross-sectoral groups (academic-private sector institutional coalitions, and international linkages). Within these groups, some projects have been chosen to host a quasi-experimental evaluation of the award activity: this is called out in the logic model. The subset of enhanced evaluations is shown as feeding directly into the program outcome measures.

Within the set of awarded projects, each one will have administrative conditions that will influence their operations and impact their success. These are listed in the box “Project-level Administration Conditions. While not all projects will experience all of the factors in the box, each one may come into play, and certainly will be important to consider in the process of conducting evaluation of the program as a whole, as well as on the award-level evaluations. In addition to these administrative factors, SRI International will be collecting data at the project level. These measures will also feed into the evaluation. This feature is depicted at the level of internal NSF actions since these are raw data that will contribute to the operation and evaluation of the projects and the program as a whole.

Program Goals

The CPATH Program goals derive from all the factors discussed above--external influences, NSF actions, project characteristics and administrative conditions. The goals
have been evolving as the CPATH Program has unfolded within the computer science education community and within NSF. The goals listed in the logic model are those that have remained as goals since CPATH has developed. We expect these goals to change as the program continues to evolve. As currently construed, the goals feed directly into program outcomes, but the outcomes will be different based upon whether the goals are targeted at institutional transformation, community building, or some other aspect of computing education.

Program Outcomes and Measures: Transformation or Community Building

Model 1 diverges into A and B at the point of whether the model is depicting institutional transformation or community building. The diversion at this point is due to the fact that these two tracks require different measures as we try to understand and evaluate the extent to which the Program is meeting its goals. The goals are very similar in each track, but the way the outcomes will be measured will be different. For institutional transformation in 1A, the measures focus on the structure of departments and institutions as well as the extent to which curricula extends beyond computer science or engineering departments. For community building in 1B, the measures focus on social networks and knowledge diffusion through newly-formed connections and groups. The measures listed in the SRI logic models focus on quantitative measures; qualitative input will be derived from site visit reports combined with reports based on the efforts of on-site evaluators or quasi-experimental evaluators who are working at the project level.

Glossary of Terms

This report contains a glossary of evaluation terms derived from the literature. These terms are chosen based on the extent to which they are relevant to the CPATH Project. The evaluation community has been attentive to issues of quality and the need for well designed randomized controlled trials or other quantitative measures wherever possible. Recently, the membership of the American Evaluation Association voted to revise The Guiding Principles for Evaluation (Shadish, Newman, Scheirer, & Wye, 1995). The Joint Committee on Standards for Educational Evaluation recognized a similar concern regarding the standards and criteria for evaluation practice and anticipated a revision of their program evaluation standards (Joint Committee, 1994). The evaluation community’s focus on issues of quality is also evidenced by the popularity of two Web sites: the OERL Web site which provides examples of sound evaluation reports, plans, and instruments designed by SRI International; and the Western Michigan University Web site which provides checklists to assess the quality of an evaluation. The glossary of evaluation terms has been compiled from these sources to support rigorous evaluation; the compilation of terms and definitions is included in this report, and it is on the project website at cpathmonitor.org.

Additions to the glossary of terms are welcome; they can be submitted on the website.
Model 1A. Annual Program Evaluation Logic Model for Institutional Transformation Aspects of CPATH

Precedents

Internal NSF Actions

Program Goals

Program Outcomes - Transformation

Measures

![Diagram of Model 1A. Annual Program Evaluation Logic Model for Institutional Transformation Aspects of CPATH]
NSF Context
Foundation Strategic Plan
IT becoming part of many sciences
Cross-NSF cooperation
Spurred by ACC

National Context
Computing transforming society
Applications more complex
U.S. leadership essential
Drop in computer education enrollment
National invigoration viewed as essential

NSF CISE Actions
Community awareness workshops
Goal development, evolution
Program design, funding
Proposal solicitation, peer review, feedback
Award distribution
Post-award assistance

Solicitation & Award Characteristics
Community Building Concepts
Transformative Implementation
Conceptual Development and Planning
Geographic Spread, Institutional Diversity

Awardee Characteristics
Single Institution
Multiple Institutions
Cross-sectoral Projects
International Linkages

SRI Project Monitoring and Site Visits

Model 1B. Annual Program Evaluation Logic Model for Community Building Aspects of CPATH

Knowledge of computing broadened, deepened
Models replicated across institutions
Computational thinking integrated across disciplines
Sustained focus on innovation
Develop alternative models

Change institutional structure
Attract different student cohort
Contribute to diverse industries

Deepen knowledge of and use of computing
Models replicated across institutions
Computational thinking integrated across disciplines
Sustained focus on innovation
Develop alternative models

Count disciplines involved at faculty level compared to bibliometric expectations/observations
Measure impact strength of articles from chosen departments compared to non-participants
Network analysis of collaborations across disciplines, institutions
Counterfactual measure of intervention impact based upon pre-program data
Identify external support connected to the interest generated by award
Web hits and web links to project website compared to similar program
Number and nature of connections among departments compared to pre-project activity
Test model in focus groups or scenario plans
Network measures of changes in connections across sectors

Scientific, technical advances
Changes to NSF budget, & to Federal, state, local education budgets
Changes in industry needs & support for education
Availability of student support
Public and media attitudes ("geeks")

External Influences

National Context
Computing transforming society
Applications more complex
U.S. leadership essential
Drop in computer education enrollment
National invigoration viewed as essential

NSF Context
Foundation Strategic Plan
IT becoming part of many sciences
Cross-N NSF cooperation
Spurred by ACC

External Influences
Scientific, technical advances
Changes to NSF budget, & to Federal, state, local education budgets
Changes in industry needs & to support for education
Availability of student support
Public and media attitudes ("geeks")

Precedents
Internal NSF Actions
Program Goals
Program Outcomes
Measures
External Influences
Scientific, technical advances
Changes to NSF budget, & to Federal, state, local education budgets
Changes in industry needs & to support for education
Availability of student support
Public and media attitudes ("geeks")

National Context
Computing transforming society
Applications more complex
U.S. leadership essential
Drop in computer education enrollment
National reform viewed as essential

NSF Context
Foundation Strategic Plan
IT becoming part of many sciences
Cross-NSF cooperation
Spurred by ACC

NSF CISE Actions
Community awareness workshops
Goal development, evolution
Program design, funding
Proposal solicitation, peer review, feedback
Award distribution
Post-award assistance

CPATH Program Logic Model

Solicitation & Award Characteristics
Community Building
Transformative Implementation
Conceptual Development and Planning
Geographic, type diversity

Awardee Characteristics
Single Institution
Multiple Institutions
Cross-sectoral Projects
International Linkages

Quasi-experimental evaluation projects

Project-level Administration Activities
Evaluation plan, data collection
Project timeframe, milestones
Data tracking, management
Contact of faculty, administration
Access of equipment, supplies
Meeting space
Establishment of collaborative links
Amendments to objectives, staffing, partners
Unforeseen obstacles to implementation
Unforeseen opportunities for expansion

CB Track
Operations
Outputs
Outcomes
Measures

CDP Subtrack
Operations
Outputs
Outcomes
Measures

TI Subtrack
Operations
Outputs
Outcomes
Measures
CPATH Program Logic Model Continued
(CB Track)

- **Operations**
  - Producing workshops, forums, competitions, online collaborations
  - Applying teaching methods from other fields
  - Developing new curricula, degree programs
  - Enlisting non-traditional students
  - Engaging new stakeholders

- **Outputs**
  - New teaching models/methods/materials
  - New partnerships (industry, international)
  - New faculty collaborations
  - Workforce needs assessments
  - Online education resources
  - TI proposals

- **Outcomes**
  - More students attracted to UCE
  - Students better prepared for work, school
  - Faculty teach across disciplinary lines
  - More, better resources available on line
  - Collaborations continue, expand
  - TI proposals implemented

- **Measures**
  - Student demographics, attitudes change
  - Faculty, employers more satisfied
  - Workshop, forum recognized as important contribution to field
  - Partners value collaborations
CPATH Program Logic Model Continued
(TI Subtrack)

**Operations**
- Create or streamline program pathways
- Create changes in culture, practice
- Foster faculty collaborations
- Create new educational materials
- Evaluate models being implemented
- Extend existing models, approaches
- Serve as national resource

**Outputs**
- New materials, pathways, practices
- Organizational, cultural changes
- Incorporation of computational thinking

**Outcomes**
- Awardee institutions make systemic changes
- Models, materials, etc. available to other institutions

**Measures**
- Changes sustained at awardee institutions
- Models, etc. adopted, adapted by others
CPATH Program Logic Model Continued
(CDP Subtrack)

- Operations:
  - Convening workshops
  - Designing concepts
  - Evaluating programs/models
  - Gaining faculty support
  - Gaining organizational support
  - Developing implementation plans

- Outputs:
  - Full-scale TI proposals
  - Implementation plans
  - New collaborations across disciplines, institutions

- Outcomes:
  - TI proposal funded, implemented
  - If not, some changes adopted

- Measures:
  - Proposal is well formulated
  - Implementation plan sound
  - Institutional commitments realizable
### Compiled Glossary of Terms Related to Evaluation and Monitoring

<table>
<thead>
<tr>
<th><strong>Term</strong></th>
<th><strong>Definition</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analysis of variance</strong></td>
<td>A widely-used statistical inference technique, based on comparing the variance between samples with the variance within samples. This can tell us whether there is any systematic difference between samples that needs to be explained.</td>
</tr>
<tr>
<td><strong>Analytical agenda</strong></td>
<td>A logical structure imposed on the different questions to be asked in an evaluation. This serves to transform the general, often vague, questions which those requesting the evaluation have in mind into questions which are precise enough to be manageable by evaluation research methods. Once the analytical agenda has been drawn up, those responsible for commissioning the evaluation have to ask whether the intervention is indeed evaluable in terms of this analytical agenda.</td>
</tr>
<tr>
<td><strong>Audit</strong></td>
<td>A control function, which is primarily concerned with verifying the legality and regularity of the implementation of resources in a program. Audit has traditionally covered areas such as the verification of financial records (financial audit).</td>
</tr>
<tr>
<td><strong>Baseline study</strong></td>
<td>The collection and analysis of data regarding a target audience or situation prior to intervention. Generally, baseline data are collected in order to provide a point of comparison for an evaluation.</td>
</tr>
<tr>
<td><strong>Before-and-after design</strong></td>
<td>An example of a quasi-experimental design in which one simply compares the relevant state of the world after the intervention with its state beforehand and attributes any difference to the effects of the intervention. A particular weakness of this design is the possibility that something else besides the intervention accounts for all or part of the observed difference over time. See also control group, counterfactual situation, evaluation design, internal validity, intervention logic, quasi-experimental designs, and program group.</td>
</tr>
<tr>
<td><strong>Benchmarks</strong></td>
<td>Standards by which the performance of an intervention can be assessed in a non-arbitrary fashion. An obvious way of deriving benchmarks would be to examine the intervention's objectives.</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>as expressed by expected outputs, results and outcomes. Ideally, benchmarks should allow us to compare the performance of an intervention with that of other policy instruments in the same field of action or in a related one. See also general objectives, indicator, intervention, objectives, operational objectives, outcomes, outputs, results, specific objectives.</td>
<td></td>
</tr>
<tr>
<td>Case studies</td>
<td>A data collection technique involving the examination of a limited number of specific cases or projects which the evaluator anticipates will be revealing about the program as a whole. Case studies tend to be appropriate where it is extremely difficult to choose a sample large enough to be statistically generalizable to the population as a whole; where generalization is not important; where in-depth, usually descriptive data is required; and where the cases or projects to be studied are likely to be quite complex. See also case study designs, data collection.</td>
</tr>
<tr>
<td>Case study designs</td>
<td>A class of evaluation designs in the descriptive rather than the causal approach. It is often the case that an evaluation design will be based on an in-depth study of one or more specific cases or situations. See also case studies, evaluation design.</td>
</tr>
<tr>
<td>Comparative change design</td>
<td>An example of a quasi-experimental design in which any known or recognizable difference between the program and control groups is taken into account in the statistical analysis. The problems with this design are, firstly, that there may be some other factor which explains some or all of the variation in the intervention and in the observed effects, and, secondly, that there may be initial differences between the program and control groups which have an influence on observed effects and which can therefore become confounded with the influence of the program on these effects (selection bias). See also control group, counterfactual situation, evaluation design, internal validity, intervention logic, quasi-experimental designs, program group, selection bias.</td>
</tr>
</tbody>
</table>
| Control group            | A group of subjects which have not been exposed to an intervention. The control group should resemble the program group (the subjects which
have been exposed to the intervention), so that systematic differences between the two groups may be attributed to the effects of the intervention once other plausible alternative hypotheses have been eliminated or discounted. See also counterfactual situation, evaluation design, intervention logic, program group.

| **Cost-benefit analysis** | A judgmental technique in which a researcher compares all social and private costs and benefits of a program with a view to determining whether the benefits exceed the costs, and if so by how much. Social costs and social benefits usually have to be measured by some indirect means and converted into monetary values so that a comparison can be made with private costs and benefits. Furthermore, it may not be appropriate to use prevailing market prices. Consider a situation of very high unemployment. In this case, the real cost of labor may be much lower than the prevailing market wage. The opportunity cost (the next best use of the otherwise unemployed workers had the project not gone ahead) is lower than the prevailing wage rate, and this low opportunity cost has to be represented by a shadow price which has to be derived somehow. See also cost-effectiveness analysis. |
| **Cost-effectiveness analysis** | A judgmental technique in which the researcher quantifies the costs and benefits associated with a program on the basis of the same principles which apply to cost-benefit analysis, but there is no requirement to transfer benefits into common monetary units. See also cost-benefit analysis, effectiveness. |
| **Counterfactual analysis** | The situation which would have arisen had the intervention not taken place. In order to derive the counterfactual situation we need an evaluation design. Except for the theoretical case of the ideal experimental design, we can never know the counterfactual situation with certainty. Real world evaluation designs tend to be based on an estimate of the counterfactual derived either from comparing subjects who were exposed to an intervention with a comparison group who were not exposed, or from examining subjects before and after exposure. See also control group, evaluation design, ideal experimental design. |
| **Criterion-population design** | An example of a quasi-experimental design, which attempts to improve on the comparative change design. In the latter, the program and control groups are two distinct groups drawn from a hypothetical larger population. In the criterion-population design, however, the hypothetical population is identified and used for the comparison group. In this case, the possibility of selection bias is confined to just one group - the program group. This design is particularly appropriate where the evaluator cannot easily create a control group but does have access to information about the larger population from which the program group is drawn. See also control group, comparative change design, counterfactual situation, evaluation design, internal validity, intervention logic, quasi-experimental designs, program group, selection bias. |
| **Data** | Known facts which can be used as a basis for inference. Subjective data involve personal feelings, attitudes and perceptions; objective data relate to observable facts. Quantitative data involve numerical observations; qualitative data are non-numerical and related to categories. Longitudinal data are collected over time; cross-sectional data are collected from the same point in time, but from a variety of different geographical areas, etc. Primary data are taken directly from original sources or collected first hand; secondary data have undergone extensive manipulation and interpretation. See also data analysis, data collection. |
| **Data analysis** | The main techniques used to interpret information about an intervention for use in an evaluation are statistical analysis, the use of models, non-statistical analysis and judgment techniques, such as cost-benefit analysis, cost-effectiveness analysis and multi-criteria analysis. See also cost-benefit analysis, cost-effectiveness analysis, data collection, models, multi-criteria analysis, non-statistical analysis, statistical analysis. |
| **Data collection** | The main techniques used to gather information about an intervention for use in an evaluation are surveys, case studies, natural observations, expert
opinion, reviews of program documents and literature reviews. See also case studies, data analysis, evaluation design, expert opinion, literature reviews, natural observations, program document reviews, and surveys.

| **Deadweight** | Deadweight is defined as effects which would have arisen even if the intervention had not taken place. Deadweight usually arises as a result of inadequate delivery mechanisms which fail to target the intervention's intended beneficiaries sufficiently well. As a result, other individuals and groups who are not included in the target population end up as recipients of benefits produced by the intervention. Deadweight is really a special case of program inefficiency. See also delivery mechanisms, efficiency, target population. |
| **Delivery mechanisms** | The organizational arrangements which provide the goods and services funded by the intervention to its intended beneficiaries, i.e. its target population. See also target population. |
| **Delphi technique** | A technique which can be used to systematize expert opinions. Experts are consulted separately in a number of different rounds. In each successive round, each individual is told the views of the other experts in the previous round. This technique can be used to arrive at a consensus, or at least to reduce disagreements. |
| **Displacement** | Displacement and substitution are two closely related terms which are used to describe situations where the effects of an intervention on a particular individual, group or area are only realized at the extent of other individuals, groups or areas. Consider, for example, the case of a program to provide employment subsidies. In a firm which benefits from this program, subsidized workers may take the place of unsubsidized workers who would otherwise have been employed by that firm. This is known as substitution. Alternatively, a firm benefiting from the employment subsidies may win business from other firms which do not participate in the scheme. Thus, the jobs created in the participating firm may be partly or wholly offset by job losses in other firms. This is known as displacement. |
| **Double-loop learning** | A type of feedback, in which the information |
| Compiled by an evaluation is used to call into question the very existence of an intervention or to bring about major changes in its basic orientations. Double-loop learning is almost always the result of summative evaluations. It is of key importance in focusing the activities of the European Union towards meeting the evolving needs of its citizens. See also feedback, formative evaluation, single-loop learning, summative evaluation. |

| **Effectiveness** | To what extent have the intervention's impacts contributed to achieving its specific and general objectives? See also cost-effectiveness analysis, general objectives, impacts, intervention logic, objectives, outcomes, results, specific objectives. |

| **Efficiency** | How economically have an intervention's inputs been converted into outputs and results? See also inputs, intervention logic, outputs, results. |

| **Evaluability** | The issue of whether or not the questions raised by a given analytical agenda for an evaluation are at all answerable by an evaluator using appropriate research methods. To know whether the questions can be answered with an acceptable degree of credibility, it is often advisable to perform an evaluability assessment. If an intervention is not evaluable in terms of this analytical agenda (e.g. because adequate data are not yet available), this can lead to a decision to postpone the evaluation or to draw up a new, more realistic analytical agenda. See also analytical agenda, evaluability assessment, evaluation project. |

| **Evaluability assessment** | An attempt to determine whether or not the questions raised by a given analytical agenda for an evaluation are at all answerable by an evaluator using appropriate research methods. See also analytical agenda, evaluability, and evaluation project. |

| **Evaluation** | An in-depth study using recognized procedures to assess a component of an intervention that compares what was expected to what was observed. |

| **Evaluation design** | A model which is used to describe an intervention and provide evidence on the effects which may be attributable to it. Evaluation designs are either causal or descriptive in nature. A given design |
should lead to the choice of one or more data analysis and collection techniques. See also counterfactual situation, data analysis, data collection, ideal experimental design, and intervention logic.

<p>| Evaluation project | A sequence of logical steps starting out from the formulation of problems and interests motivating the evaluation to arrive at a series of questions that can be addressed in an analytically acceptable way. Typically, the commissioner establishes an evaluation plan that sets out a framework in which the evaluation is to be conducted and then chooses the evaluator. |
| Evaluation report | The end product of an evaluation, the evaluation report should follow a logical structure and meet the needs of the evaluation sponsors and the principal stakeholders. Evaluation reports should include an executive summary. The structure of the report is usually specified by the sponsors in the terms of reference. See dissemination, evaluation sponsors, executive summary, reporting, stakeholders, terms of reference. |
| Evaluation sponsors | The department or agency responsible for commissioning the evaluation. See also management structure, organizational structure, stakeholders, steering group, and terms of reference. |
| Ex ante evaluation | An evaluation conducted before the implementation of an intervention. Also referred to as an &quot;appraisal&quot;. See also evaluation, ex post evaluation, intermediate evaluation. |
| Ex post evaluation | An evaluation conducted either on or after completion of an intervention. See also evaluation, ex ante evaluation, intermediate evaluation. |
| Ex post facto design | An example of a descriptive design, which can be used where the evaluator cannot select who is to be exposed to the program, and to what degree. These designs have been used to examine interventions with universal coverage. See also control group, counterfactual situation, evaluation design, intervention logic, program group. |
| Executive summary | It is likely that only a small proportion of the target audience will read the full evaluation report. It is therefore advisable to produce a well-written executive summary of no more than five |</p>
<table>
<thead>
<tr>
<th>Concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert opinion</td>
<td>A data collection technique, similar to a survey, which relies on the necessarily subjective views of experts in a particular field. It is not recommended to rely on expert opinion as a sole data source, for example, because of problems with so-called &quot;chatty bias&quot;. See also chatty bias, data collection, Delphi technique, surveys.</td>
</tr>
<tr>
<td>External evaluation</td>
<td>An evaluation which is performed by persons outside the organization responsible for the intervention itself. See also evaluation, internal evaluation.</td>
</tr>
<tr>
<td>External validity</td>
<td>The confidence one can have about whether or not one's conclusions about the intervention can be generalized to fit circumstances, times, people, and so on, other than those of the intervention itself. A threat to external validity is an objection that the evaluation design does not allow causal inference about the intervention to be generalized to different times, places or subjects to those examined in the evaluation. See also evaluation design, internal validity, intervention logic.</td>
</tr>
<tr>
<td>Feedback</td>
<td>The process by which the information compiled by an evaluation is used by decision-makers to either change the way in which an intervention is implemented, or to bring about a more fundamental change in the basic orientations of the intervention, including calling into question its very existence. See also double-loop learning, single-loop learning.</td>
</tr>
<tr>
<td>Formative evaluation</td>
<td>An evaluation concerned with examining ways of improving and enhancing the implementation and management of interventions. Formative evaluations tend to be conducted for the benefit of those managing the intervention with the intention of improving their work. See also evaluation, summative evaluation.</td>
</tr>
<tr>
<td>General objectives</td>
<td>The desired effects of an intervention expressed in terms of outcomes, i.e. the longer-term impact of the intervention on society (e.g. to reduce unemployment among the long-term unemployed). See also intervention logic, objectives, operational objectives, outcomes,</td>
</tr>
</tbody>
</table>
### Hawthorne effect

The term "Hawthorne effect" is used to explain situations where an experiment cannot be trusted because the very fact that the experiment is taking place is influencing the results obtained. This reminds us that program staff and beneficiaries can behave quite differently from their normal patterns if they know that they are being observed.

### Ideal experimental design

A theoretical way of deriving the counterfactual situation, and hence the net impact of an intervention. It involves comparing two groups which are identical in all respects except one: exposure to the intervention. Differences between the group which has been exposed (the program group) and the group which has not (the control group) are then attributable to the intervention. In the real world, this design does not exist since we can never be absolutely certain that the two groups are identical in all other respects. The potential non-equivalence of the two groups weakens the validity of any causal inference about the intervention. A number of real world evaluation designs are available which each have their own strengths and weaknesses. See also control group, counterfactual situation, evaluation design, intervention logic, program group, quasi-experimental designs, and true experimental designs.

### Impacts

A general term used to describe the effects of a program on society. Impacts can be either positive or negative and foreseen or unforeseen. Initial impacts are called results or outputs, while longer-term impacts are called outcomes.
<table>
<thead>
<tr>
<th><strong>Impact indicators</strong></th>
<th>Those specific measures that assess the extent to which a project accomplished its stated goals and objectives. Also called impact, outcome, or summative evaluation. Impact evaluations focus on the end results of projects.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indicator</strong></td>
<td>A characteristic or attribute which can be measured to assess an intervention in terms of its outputs or impacts. Output indicators are normally straightforward. Impact indicators may be more difficult to derive, and it is often appropriate to rely on indirect indicators as proxies. Indicators can be either quantitative or qualitative. The term &quot;performance indicators&quot; is also used. See also benchmarks, general objectives, impacts, operational objectives, outputs, specific objectives.</td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td>The human and financial resources involved in the implementation of an intervention. See also intervention, intervention logic.</td>
</tr>
<tr>
<td><strong>Intermediate evaluation</strong></td>
<td>An evaluation conducted during the implementation of an intervention. See also evaluation, ex ante evaluation, ex post evaluation.</td>
</tr>
<tr>
<td><strong>Internal evaluation</strong></td>
<td>An evaluation which is performed by members of the organization responsible for the intervention itself. See also self evaluation, external evaluation.</td>
</tr>
<tr>
<td><strong>Internal validity</strong></td>
<td>The confidence one can have in one's conclusions about what the intervention actually did accomplish. A threat to internal validity is an objection that the evaluation design allows the causal link between the intervention and the observed effects to remain uncertain. It may be thought of as a question of the following nature: could not something else besides the intervention account for the difference between the situation after the intervention and the counterfactual? See also counterfactual situation, evaluation design, external validity, intervention, intervention logic, selection bias.</td>
</tr>
<tr>
<td><strong>Interrupted time-series design</strong></td>
<td>An example of a quasi-experimental design. It involves obtaining several measurements over time both before and after exposure to a program in order to create a time series of observations. It is an improvement on the before-and-after design. See also before-and-after design, control group, counterfactual situation, evaluation design,</td>
</tr>
<tr>
<td><strong>Intervention logic</strong></td>
<td>The conceptual link from an intervention's inputs to the production of its outputs and, subsequently, to its impacts on society in terms of results and outcomes. The examination of the program's intervention logic will be of central importance in most evaluations. The evaluator needs to ask how the program achieves its specific objectives, and how do the specific objectives contribute to the attainment of the general objectives? The terms &quot;theory of action&quot;, &quot;program logic&quot; and &quot;program theory&quot; are sometimes used to mean more or less the same thing. See also general objectives, impacts, inputs, intervention, objectives, operational objectives, outcomes, outputs, results, specific objectives.</td>
</tr>
<tr>
<td><strong>Literature reviews</strong></td>
<td>A data collection technique which enables the evaluator to make the best use of previous work in the field under investigation and hence to learn from the experiences and findings of those who have carried out similar or related work in the past. There are two types of documents that can be used in a literature search. Firstly, there are published papers, reports and books prepared by academics, experts and official organizations. Secondly, there are specific studies in the area, including past evaluations. See also data collection, research synthesis.</td>
</tr>
<tr>
<td><strong>Management structure</strong></td>
<td>A hierarchical structure which allows for overall management of an evaluation, and, in particular, the evaluation project. As a minimum, such a management structure should involve the program management (usually the same as the evaluation sponsors) and the unit, sector, or official inside the same Department/Directorate in charge of evaluation. However, for an evaluation to be successful, it may be necessary to widen the management structure and create a steering group. See also evaluation project, evaluation sponsors, organizational structure, stakeholders, steering group.</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>The most commonly used descriptive statistic, it tells us the average of a set of values. See also standard deviation, statistical analysis.</td>
</tr>
<tr>
<td><strong>Models</strong></td>
<td>There are various different models which seek to</td>
</tr>
</tbody>
</table>
represent how an intervention changes important socio-economic variables. Such models are normally taken from previous research. The main types of models are: (i) input-output models, which allow a researcher to systematically examine the linkages between the different parts of an economy, as the inputs of one industry can be thought of as the outputs of other industries; (ii) microeconomic models, which are designed to examine the behavior of households and firms in specific industries and markets using equations which represent the supply and demand functions for a particular good or service; (iii) macroeconomic models, which are used to model the behavior of the economy as a whole and the evolution of important macroeconomic variables (such as inflation, employment, growth and the trade balance) over time; and, (iv) statistical models, which are frequently used to examine relationships between specific program effects. See also data analysis, statistical analysis.

<table>
<thead>
<tr>
<th>Monitoring</th>
<th>The continuous process of examining the delivery of program outputs to intended beneficiaries, which is carried out during the execution of a program with the intention of immediately correcting any deviation from operational objectives. Also measures the extent to which members of the target audience are exposed to project components. Also measures the audience’s response to the project, including their feelings and perceptions about it and whether they attended and understood an activity or message.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multi-criteria analysis</td>
<td>A decision-making tool which can be adapted to form judgments about interventions. Multi-criteria analysis allows us to formulate judgments on the basis of multiple criteria, which may not have a common scaling and which may differ in relative importance.</td>
</tr>
<tr>
<td>Natural observations</td>
<td>A data collection technique in which the evaluator makes on-site visits to locations where the intervention is in operation and directly observes what is happening. Observational data can be used to describe the setting of the intervention, the activities which take place in the setting, the individuals who participate in these activities (who may or may not be aware that they are being</td>
</tr>
</tbody>
</table>
observed), and the meaning of these activities to the individuals. This form of data collection is particularly vulnerable to the Hawthorne effect. See also data collection, Hawthorne effect.

<p>| Needs | The socio-economic problems which an intervention aims to address, expressed from the point of view of its target population. For example, the need to improve job opportunities for long-term unemployed workers who may suffer from a lack of relevant skills. See also objectives, target population. |
| Non-statistical analysis | A general term used to describe the analysis of mainly qualitative data which is typically used in conjunction with statistical analysis (of either qualitative or quantitative data). Usually, this includes an assessment of the reliability of any findings derived from such methods. See also data, data analysis, statistical analysis. |
| Objectives | The desired effects of an intervention. See also general objectives, needs, operational objectives, specific objectives. |
| Operational objectives | The desired effects of an intervention expressed in terms of outputs, i.e. the goods and services produced by an intervention (e.g. to provide professional training courses to the long-term unemployed). See also general objectives, intervention, intervention logic, objectives, outputs, specific objectives. |
| Organizational structure | Specifying the evaluation's organizational structure, which is usually done in the terms of reference, involves delineating the role of different actors (especially important if the evaluation task is to be divided among different evaluators - for example, between internal and external evaluators), establishing reporting responsibilities (including, where appropriate, contact with evaluation steering groups, program managers, other Commission services and Member State administrations) and identifying the procedure to be followed to disseminate and use the evaluation. See also dissemination, evaluation project, external evaluation, feedback, internal evaluation, management structure, stakeholders, steering group, terms of reference. |
| Outcomes | The longer-term impact, usually expressed in terms of broad socio-economic consequences, |</p>
<table>
<thead>
<tr>
<th><strong>Outputs</strong></th>
<th>The goods and services produced by an intervention (e.g. training courses for the long-term unemployed). See also intervention, intervention logic, operational objectives.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance audit</strong></td>
<td>Conceptually closer to evaluation than traditional audit, performance audit is strongly concerned with questions of efficiency (of an intervention's direct outputs) and good management. Performance audit and evaluation share the same aim of improving the quality of programs, but evaluation goes much further. It also looks at issues such as sustainability, relevance and the longer-term consequences of a program. See also audit, evaluation.</td>
</tr>
<tr>
<td><strong>Policy</strong></td>
<td>A set of activities, which may differ in type and have different direct beneficiaries, directed towards common general objectives. Policies are not delimited in terms of time schedule or budget. See also general objectives, intervention, program, project.</td>
</tr>
<tr>
<td><strong>Population</strong></td>
<td>In statistics, the entire aggregate of individuals or subjects, from which samples can be drawn.</td>
</tr>
<tr>
<td><strong>Probability sampling</strong></td>
<td>A statistical technique used to obtain samples from a given population, whereby every unit in the population has a known, non-zero probability of being selected for inclusion in the sample. The conclusions from this type of sample can then be projected, within statistical limits of error, to the wider population. See also population, sample.</td>
</tr>
<tr>
<td><strong>Process indicators</strong></td>
<td>The specific measures of the means by which a project or policy is delivered to priority and supporting groups.</td>
</tr>
<tr>
<td><strong>Program</strong></td>
<td>A set of organized but often varied activities (a program may encompass several different projects, measures and processes) directed towards the achievement of specific objectives. Programs have a definite time schedule and budget. See also intervention, project, policy, specific objectives.</td>
</tr>
</tbody>
</table>
| **Program document reviews** | A data collection technique based on reviewing general program files, financial and administrative records and specific project
Program group | A group of subjects which have been exposed to an intervention. The program group can be compared with the control group (the subjects which have not been exposed to the intervention), in order to determine whether systematic differences between the two groups may be attributed to the effects of the intervention. See also control group, counterfactual situation, evaluation design, ideal experimental design, internal validity, intervention, intervention logic, quasi-experimental designs, true experimental designs.

Project | A single, non-divisible public intervention directed towards the attainment of operational objectives, with a fixed time schedule and a dedicated budget. See also intervention, program, policy, operational objectives.

Qualitative research | Research that collects data that appear in words rather than numbers. Useful for collecting information about feelings and impressions. Focus groups and in-depth personal interviews are common types of qualitative research.

Quantitative research | Research designed to count and measure knowledge, attitudes, beliefs, and behaviors. Yields numerical data that are analyzed statistically. Surveys are a common type of quantitative research.

Reach indicators | The specific measures that assess the degree to which priority or supporting group members are exposed to the project.

Quasi-experimental designs | A class of causal evaluation designs which take a more practical approach than is the case with true experimental designs. Control groups can still be used, but these have to be assigned through some non-random process. Alternatively, one can examine beneficiaries before and after exposure to the intervention. See also before-and-after design, comparative change design, control group, counterfactual situation, criterion-population design, evaluation design, ideal experimental design, interrupted time-series design, intervention logic, program group, true experimental designs.

Regression analysis | A statistical inference technique which can be
used to establish the significance of any correlation (association) between variables of interest, e.g. the gender of a long-term unemployed worker and the amount of time before he or she finds a new job after a training program. In regression analysis, we attempt to establish whether the variation in one variable (known as the dependent variable) can be explained in terms of the variation in one or more independent variables. The dependent variable is often quantitative, e.g. a person's income can be regressed on his educational qualifications, number of hours worked per week, age, etc. Special techniques are available, however, to deal with situations in which the dependent variable is qualitative, e.g. whether or not a person owns a car can be regressed on income, wealth, age, gender etc. See also statistical analysis.

| Relevance | To what extent are the intervention's objectives pertinent in relation to the evolving needs and priorities at both national and EU level? See also intervention, intervention logic, needs, and objectives. |
| Reporting | Reporting takes place when the evaluator transmits the evaluation report (usually in the form of a document, or else through some audio-visual presentation) to the sponsors and when they, in turn, transmit a copy (or a summary thereof) to other interested parties. See also dissemination, evaluation report, evaluation sponsors, and executive summary. |
| Research synthesis | An overview of the current state of knowledge about a socio-economic problem and about remedies through public policy, which is undertaken before an evaluation. This knowledge can be obtained from professional literature, media articles, and administrative data, monitoring reports or published statistics. Preparing a research synthesis is often helpful prior to launching an evaluation. By listing the information that is available and comparing it with the needs ensuing from the analytical agenda, the research synthesis will point to the principal information gaps which, in turn, set the data collection and analysis tasks to be undertaken by the evaluation. Reviews of |
| **literature** can also be a data collection technique in the conduct of an evaluation. See also analytical agenda, data analysis, data collection, evaluation project, literature reviews. |
|---|---|
| **Results** | The initial impact of an intervention (e.g. an improvement in the employability of the long-term unemployed through a rise in their skill level). See also impact, intervention, intervention logic, outcomes, outputs, specific objectives. |
| **Sample** | A set of individuals or items selected from a given population so that properties and parameters of the population may be estimated, or so that hypotheses about that population may be estimated. See also population, probability sampling. |
| **Scope** | The field of investigation of an evaluation. Typically, this has to be defined from an institutional (EU versus national or local level), temporal (period review) and geographical (part of the EU territory) point of view. In addition, one has to identify the key evaluation issues (relevance, efficiency, effectiveness, utility, sustainability) which will be examined. See also effectiveness, efficiency, evaluation project, relevance, sustainability, utility. |
| **Selection bias** | Could not the differences between the control group and the program group be due to initial differences in their characteristics rather than the effects of the intervention we are trying to evaluate? See also control group, counterfactual situation, evaluation design, internal validity, program group. |
| **Self evaluation** | An evaluation which is performed by members of the organization responsible for the intervention itself. See also internal evaluation, external evaluation. |
| **Single-loop learning** | A type of feedback, in which the information compiled by an evaluation is used to bring about changes in the way an intervention is implemented. Although single-loop learning is more often associated with formative evaluations, it can also arise in the case of summative evaluations. See also double-loop learning, feedback, formative evaluation, summative evaluation. |
| **Specific objectives** | The desired effects of an intervention expressed... |
in terms of results, i.e. the initial impact of the intervention on society (e.g. to improve the employability of the long-term unemployed by raising their skill level). See also general objectives, intervention, intervention logic, objectives, operational objectives, organizational structure, results, specific objectives.

Stakeholders

The various individuals and organizations who are directly and indirectly affected by the implementation and results of a given intervention, and who are likely to have an interest in its evaluation (e.g. program managers, policy-makers, the program's target population). See also evaluation sponsors, steering group, target population.

Standard deviation

A commonly used descriptive statistic, it provides a measure of dispersion for a set of values.

Statistical analysis

A commonly used data analysis technique. Statistical analysis is used often used to describe phenomena in a concise and revealing manner. This is known as descriptive statistics. It can also be used to test for relationships among variables or generalize findings to a wider population. This is known as statistical inference. See also data collection, non-statistical analysis.

Summative evaluation

An evaluation concerned with determining the essential effectiveness of programs. Summative evaluations tend to be conducted for the benefit of external actors (groups who are not directly involved in the management of a program), for reasons of accountability or to assist in the allocation of budgetary resources. See also evaluation, formative evaluation.

Surveys

A widely-used technique for collecting data from a sample drawn from a given population. Surveys are often based on probability sampling, and survey information is usually obtained through structured interviews or self-administered questionnaires. Cross-sectional surveys involve measurements made at a single point in time. Panel surveys involve measurements acquired at two or more points in time. See also data collection, population, probability sampling, sample.

Sustainability

To what extent can the program's positive impacts (as measured by its utility) be expected to last
<table>
<thead>
<tr>
<th><strong>Target population</strong></th>
<th>The intended beneficiaries of an intervention. An intervention may have more than one target population. This term should be distinguished from &quot;population&quot; in the statistical sense. See also intervention, population, stakeholders.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Terms of reference</strong></td>
<td>The terms of reference specify the work to be carried out by the evaluator, the questions to be dealt with and the time schedule. They allow the sponsors of the evaluation to define their requirements and allow the evaluator to understand clearly what is expected of the work to be undertaken (including, often, the structure of the expected evaluation report). Clearly defined terms of reference are vitally important where an evaluation is to be conducted by an external expert, and can also be of tremendous use when it is to be performed in-house. See also evaluation project, evaluation report, evaluation sponsors, external evaluation, internal evaluation, organizational structure, work plan.</td>
</tr>
<tr>
<td><strong>Thematic evaluation</strong></td>
<td>An evaluation which focuses on one or more themes which are common to several different interventions (programs or other activities), for example, effects on the environment or on small and medium-sized enterprises.</td>
</tr>
<tr>
<td><strong>True experimental designs</strong></td>
<td>The best real world approximations to the ideal experimental design, in which the evaluator tries to ensure the initial equivalence of the program and control groups by creating them beforehand through random assignment. Although causal inference based on such designs is usually very strong, true experimental designs are difficult to administer and implement. Also referred to as &quot;randomized experimental designs&quot;. See also control group, counterfactual situation, evaluation design, ideal experimental design, intervention logic, program group, quasi-experimental designs.</td>
</tr>
<tr>
<td><strong>Utility</strong></td>
<td>A comparison of a program's impacts compared with the needs of the target population. This issue is closely related to sustainability. See also impacts, intervention logic, needs, outcomes, results, sustainability, target population.</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>A descriptive statistic which provides a measure</td>
</tr>
</tbody>
</table>
of dispersion. It is obtained by squaring the standard deviation. See also analysis of variance, standard deviation, statistical analysis.

<table>
<thead>
<tr>
<th>Work plan</th>
</tr>
</thead>
<tbody>
<tr>
<td>A schema which identifies the investigations that need to be carried out by the evaluation in the light of the chief questions raised by the analytical agenda and the information gaps which have been identified. These investigations should be described in sufficient detail to provide a provisional picture of the data collection and analysis tasks lying ahead, as well as of the methodologies to be employed. In order to keep them manageable, it often proves useful to divide the various tasks to be done into different stages and to set a corresponding time-table for the delivery of the different evaluation parts. The work plan is also the appropriate place for costing the evaluation and its components. See analytical agenda, data analysis, data collection, evaluation project.</td>
</tr>
</tbody>
</table>

The definitions in this glossary have been adapted from a variety of sources, including: